Association Studies in the Human Genome

Tatiana Foroud, Ph.D.
Indiana University
Department of Medical
and Molecular Genetics

2007: The Year of Genome Wide Association Studies?

Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls

The W

Sequence variants in the autophagy gene *IRGM* and multiple other replicating loci cont susc

Miles Pa
Mark Tr
Roland

We followed up on 37 SNPs from 31 distinct loci, associated at $P < 10^{-5}$ on initial analysis of the WTCCC data set. Support for some of these markers diminished in the final WTCCC analysis after extensive data filtering[5]. We selected two markers for each locus where low linkage disequilibrium (LD) between associated SNPs in

Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes

John A Todd[1], Neil M Walker[1,9], Jason D Cooper[1,9], Deborah J Smyth[1,9], Kate Downes[1], Vincent Plagnol[1], Rebecca Bailey[1], Sergey Nejentsev[1], Sarah F Field[1], Felicity Payne[1], Christopher E Lowe[1], Jeffrey S Szeszko[1], Jason P Hafler[1], Lauren Zeitels[1], Jennie H M Yang[1], Adrian Vella[1,8], Sarah Nutland[1], Helen E Stevens[1], Helen Schuilenburg[1], Gillian Coleman[1], Meeta Maisuria[1], William Meadows[1], Luc J Smink[1], Barry Healy[1], Oliver S Burren[1], Alex A C Lam[1], Nigel R Ovington[1], James Allen[1], Ellen Adlem[1], Hin-Tak Leung[1], Chris Wallace[2], Joanna M M Howson[1], Cristian Guja[3], Constantin Ionescu-Tîrgovişte[3], Genetics of Type 1 Diabetes in Finland[4], Matthew J Simmonds[5], Joanne M Heward[5], Stephen C L Gough[5], The Wellcome Trust Case Control Consortium[6], David B Dunger[7], Linda S Wicker[1] & David G Clayton[1]

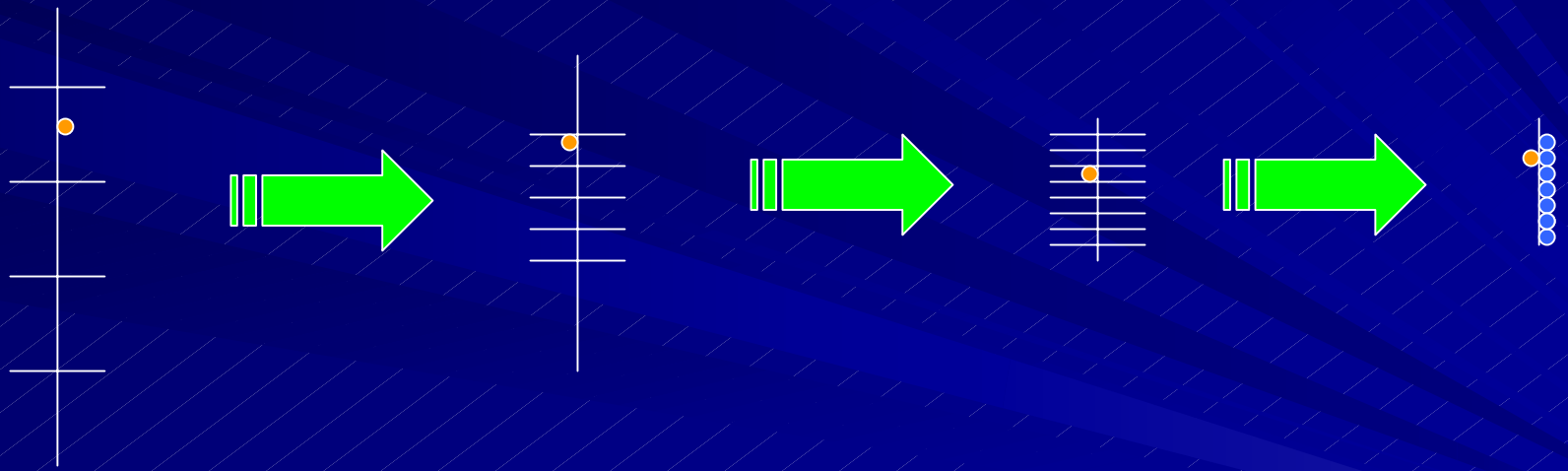
From: Teri Manolio, NHGRI

Genome Wide Association Studies (GWAS)

- Genome wide association studies are intended to provide dense coverage of the whole genome
- Dense coverage allows the detection of genes (alleles) associated with phenotypes including disease risk and therapeutic effect

Marker Development

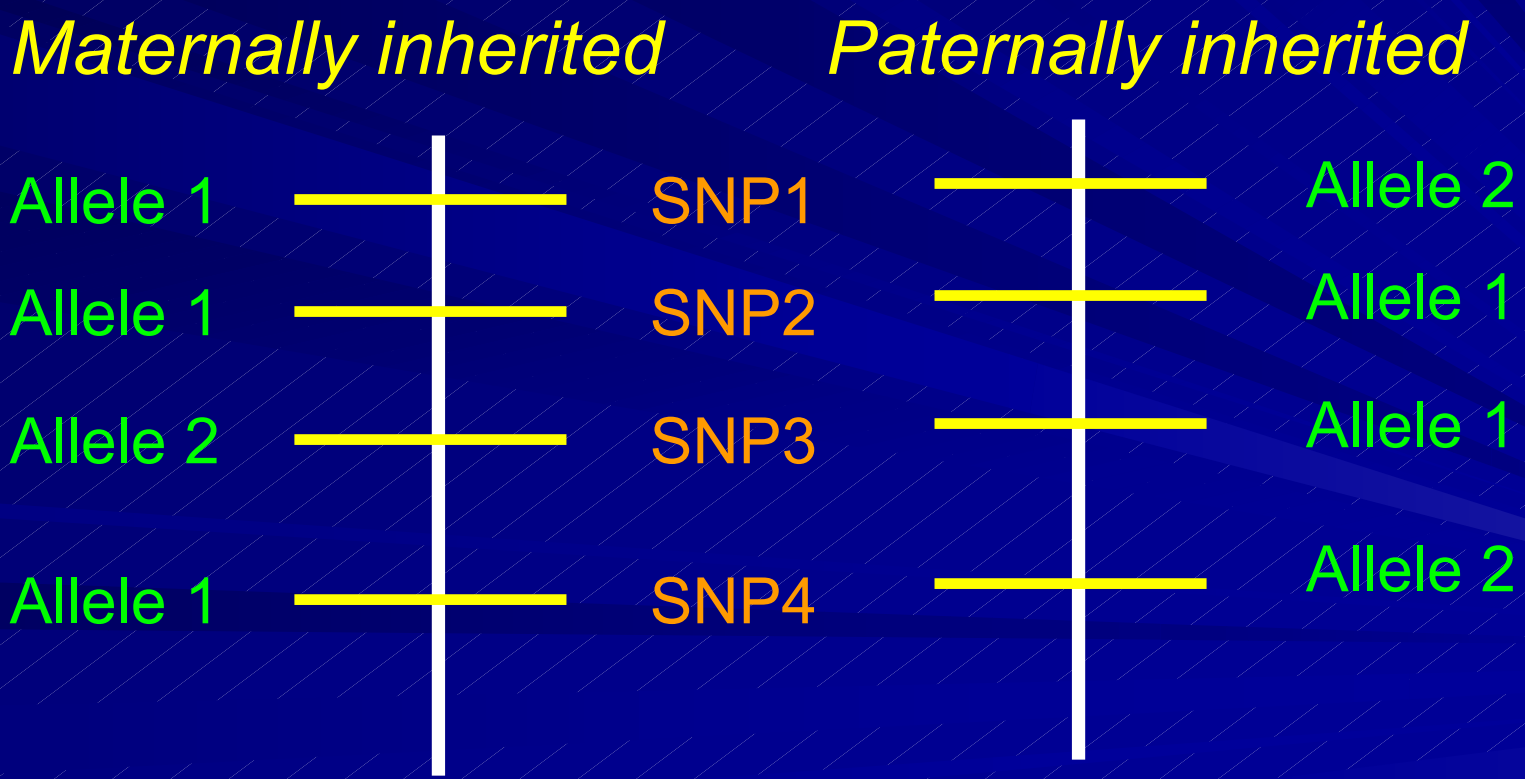
Microsatellite **SNPs** **SNPs** **SNPs**



- Increasingly denser sets of markers allow the variant or mutation of interest to be closer to the marker being tested

Haplotype

- Alleles carried on the same chromosome

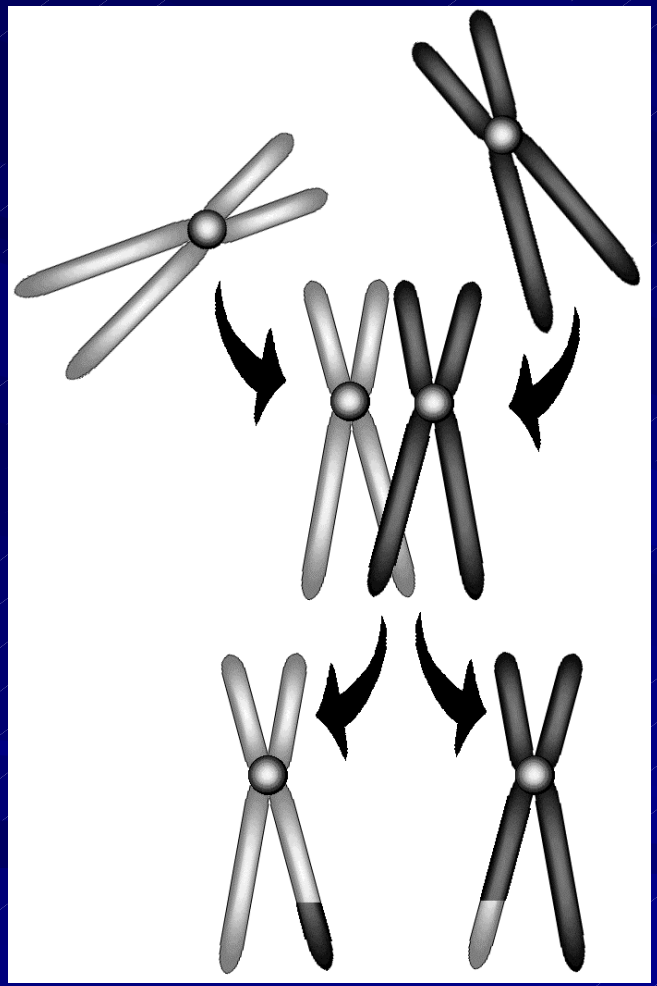


Equilibrium/Disequilibrium

	Allele 1	Allele 2
SNP1	0.4	0.6
SNP2	0.8	0.2

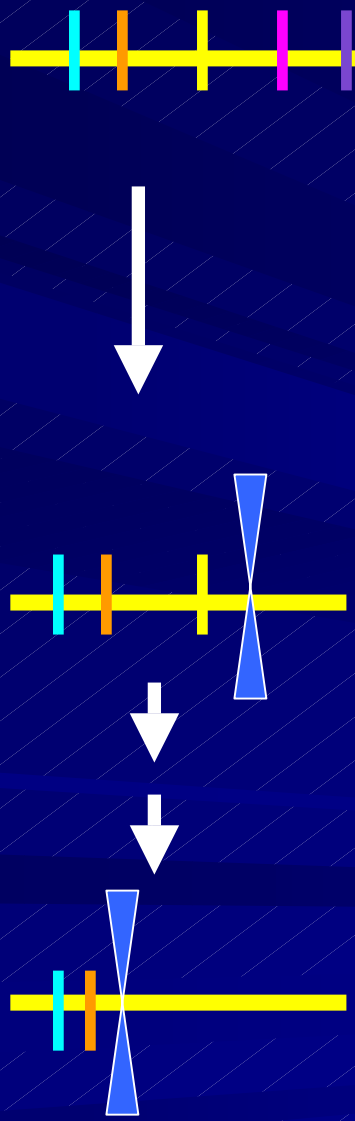
SNP1	SNP2	Equilibrium	Disequilibrium
1	1	0.4 * 0.8 = 0.32	*altered*
1	2	0.4 * 0.2 = 0.08	*altered*
2	1	0.6 * 0.8 = 0.48	*altered*
2	2	0.6 * 0.2 = 0.12	*altered*

Crossing Over / Recombination



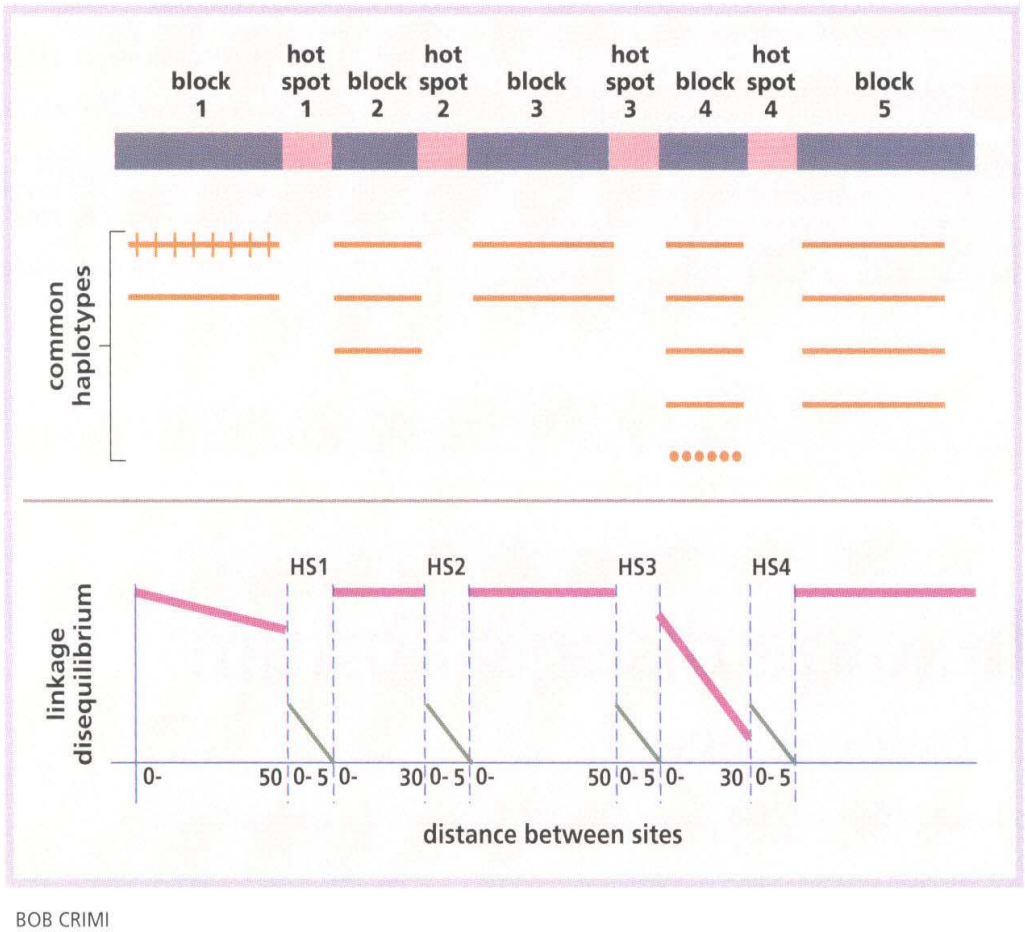
- Crossing over leads to the inheritance of new combinations of alleles at different loci
- When view marker genotypes, can only observe recombination

Recombination and Equilibrium



- Over generations, recombination occurs along the chromosome
- Recombination is less likely to occur in short physical distances
- Typically, SNPs closest together are less likely to undergo recombination

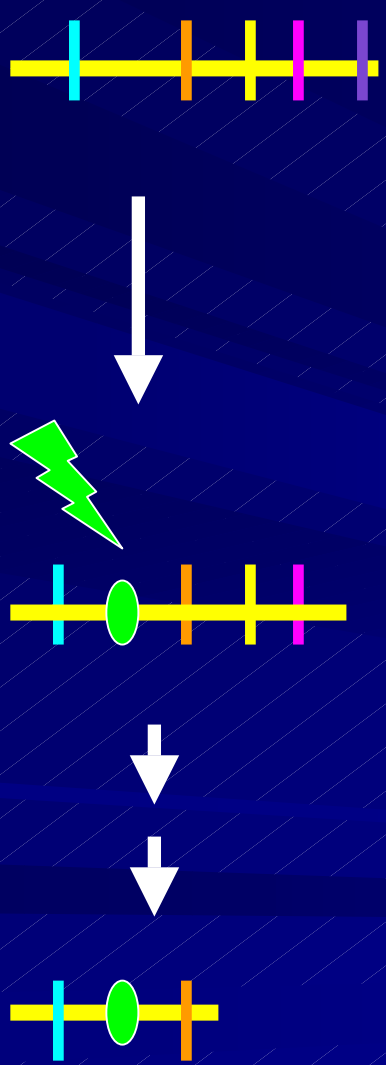
Linkage Disequilibrium



- Disequilibrium is not a linear process of decay
- Recombination hotspots exist which separate haplotype blocks
- Recombination is rare in coldspots and result in haplotype blocks

Goldstein et al. *Nature Genetics* 29: 109-111, 2001

Ancestral Variants



- Variant occurs on a particular haplotype background
- Over generations, recombination occurs, and SNPs farther from variant become in equilibrium with the variant
- SNPs closer to the variant less likely to undergo recombination, remain in linkage disequilibrium with variant

Genome Wide Association Studies

Cases

DNA inherited in blocks, so not all 10 million SNPs have to be tested

Controls

Compare frequency of SNP alleles in two groups

Compare frequency of SNP genotypes in two groups

270 Samples Included in the HapMap Project

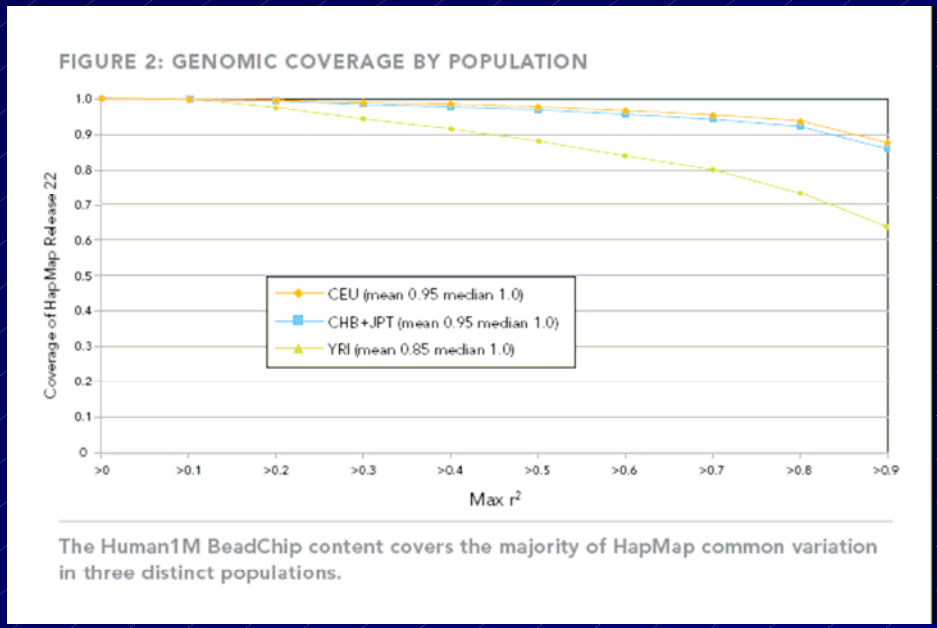
- Yoruba in Ibadan, Nigeria (YRI; 30 trios)
- Japanese in Tokyo, Japan (JPT; 45 unrelated)
- Han Chinese in Beijing, China (CHB; 45 unrelated)
- CEPH (Utah residents with ancestry from northern and western Europe) (CEU; 30 trios)

GWAS Chip Products

- New generations of GWAS chips have been rolled out rapidly
- Improvements primarily based on:
  - Number of SNPs on the chips
  - Use of HapMap information
  - SNP selection criteria and genome wide coverage
  - Inclusion of SNPs to allow comparison of copy number variation

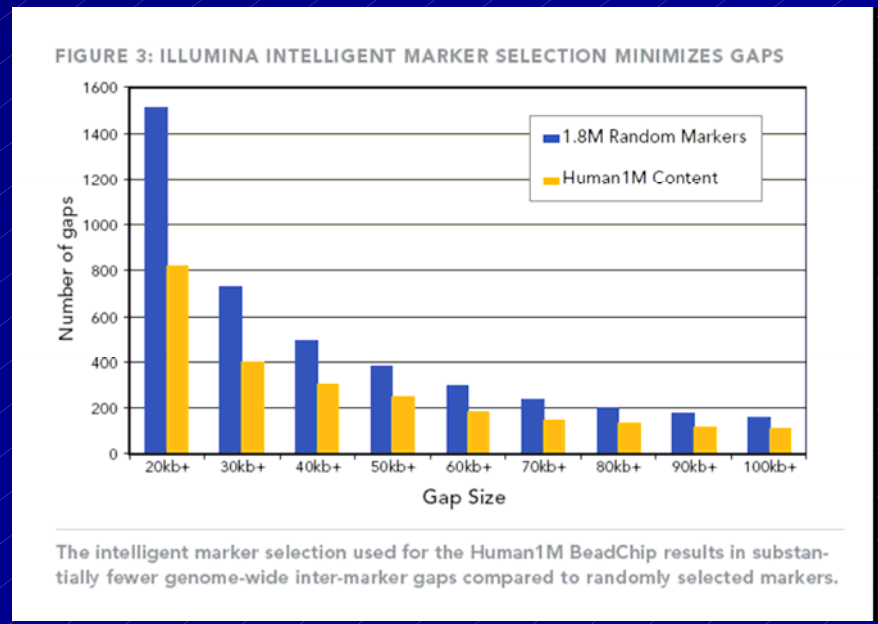
GWAS Chip Products

- Affymetrix
  - Affy 5.0 (500,000) and Affy 6.0 (1 million SNPs)
  - SNPs initially selected based on uniform physical distance
  - Newer chips select SNPs based on HapMap data
- Illumina
  - Chips with 550,000 and 1 million SNPs
  - SNPs selected based on HapMap data
  - Average: 1 common SNP (MAF ≥0.05) every 6 kb

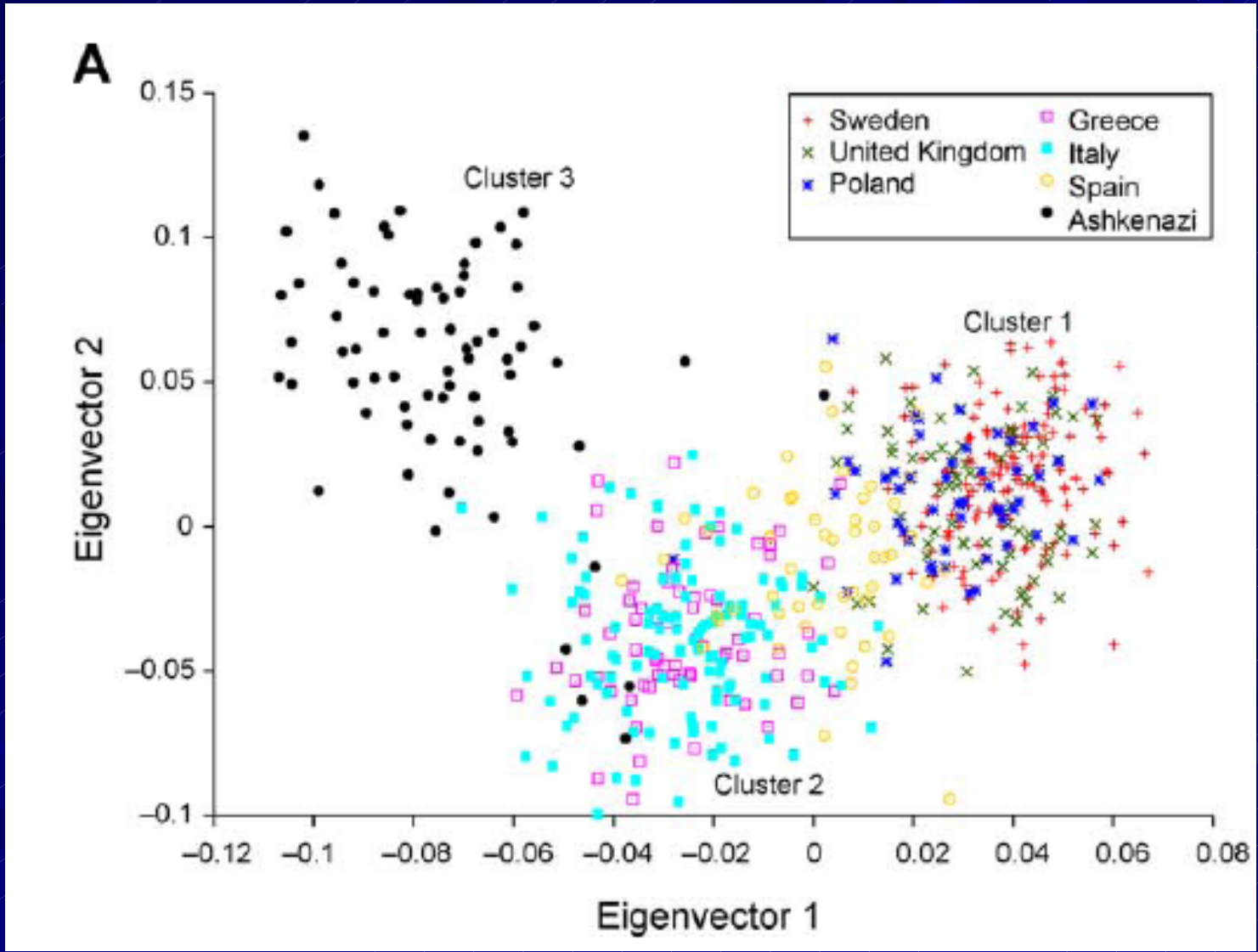


- Increasing number of SNPs provides better coverage of the entire genome for all populations
- Increase in coverage most dramatic in Yoruba

- With 1 million SNPs, gaps between markers are relatively small
- Small gaps improve the likelihood that a variant will be in linkage disequilibrium with a SNP on the chip



Sample Stratification



Price et al, PLOS Genetics, e236, 2008

What can a GWAS tell us?

- Detect genes of *large* effect in modestly sized samples (hundreds of samples)
  - Age-related Macular Degeneration and complement factor H (*CFH*) (involved in > 50% of all cases)
- Detect genes of *small* effect in very large sized samples (thousands of samples)
  - Type II Diabetes and *FTO, CDKAL1, HHEX, CDKN2B, IFG2BP2, TCF7L2, SCL30A8, KCNJ11, PPARG* (odds ratio 1.1)

What are we likely to miss with GWAS?

- This approach is best designed to detect association with *common variants*
- If many variants within a gene contribute to disease risk, power to detect association substantially reduced
- If each family has a unique variant contributing to disease risk, no power to detect association